REPORT DOCUMENTATION PAGE

DTIC ELECTE MAR 0 5 1992 D

AD-A247 032

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Program Manager Approved for public release; distribution unlimited.
IBER(S) D	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 92 0103 (2)
Daniel Kahneman Dept. of Psychology University of California, Berkeley, 94720	6b. OFFICE SYMBOL (If applicable)
	7a. NAME OF MONITORING ORGANIZATION AFOSR/NL
6c. ADDRESS (City, State, and ZIP Code) The Regents of the University of California c/o Sponsored Projects Office, M-11 Wheeler Berkeley, California 94720	7b. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR/NL	8b. OFFICE SYMBOL (If applicable)
	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-89-0206
8c. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6447	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	6912	OR	

11. TITLE (Include Security Classification)
Norms and the Perception of Events

12. PERSONAL AUTHOR(S)
Kahneman, Daniel

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Report	FROM 6-15-90 TO 9-05-91	91-09-05	32

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Social Cognition, Multiple Representations, Comparison Processes, Reference Effects

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The major effort of the research reported here has been been directed to understanding multiple representations in thinking and processes of comparison in different domains. Five distinct projects address issues of interpersonal versus intrapersonal comparisons, mental contamination, anchoring effects, topic and referent in perceptual comparison, and reference effects in choice.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr Alfred R. Fregly	22b. TELEPHONE (Include Area Code) (202) 767-5021 — 22c. OFFICE SYMBOL NL

Section 1: Comparison of Intrapersonal and Interpersonal Norms
Grant and Kahneman

This project is concerned with people's judgments of behavior in the presence of multiple frames of reference. Norm theory (Kahneman and Miller, 1986) suggests two such frames which can be used to judge an actor's behavior: the first is to locate the person's behavior relative to an interpersonal norm or frame of reference; the second is to locate the person's behavior relative to an intrapersonal norm or frame of reference. Thus, to judge the riskiness of a friend's bet at the track, the interpersonal comparison would pick out the riskiness of her bet relative to the bets of others, while the intrapersonal comparison would pick out the riskiness of this bet with respect to her previous bets. Given these two frames of reference, the question can be asked: if frame of reference is not specified, what form will peoples' judgments of behavior take? Previous research (Campbell, Fairey, & Fehr, 1986; Farkas, 1991; Hertzmen & Festinger, 1940; Levine & Green, 1984; Schul & Szyf, 1991) suggests two hypotheses: (1.) People mix the two standards when judging an actor's behavior (Mixture hypothesis), (2.) People choose one of the standards to judge the actor's behavior (Choice hypothesis). In all, four experiments have been conducted exploring these two possibilities. Each will be described in turn.

Experiment 1

An experiment was run in which subjects in three conditions made judgments of new behaviors by target actors. Two questions are addressed: (1.) do people have to choose between the standards or do they use both (mixture) in rendering their judgments of behavior? (2) which standard has a more pervasive effect upon judgment?

Method

Subjects. Seventy-seven University of California undergraduates participated in the experiment in order to fulfill a course requirement. Seven of the subjects did not follow the instructions and were deleted from the statistical analysis.

Materials. Stimulus materials consisted of nine examples. Each example centered around a particular activity -- for example, competitive sports, tips after a meal at a restaurant, performance on a math quiz, etc. -- and involved the behavior of three individuals. Three background behaviors and one target behavior were created for each person in each example; all behaviors were expressed in

quantitative terms -- batting average, number of sales, etc. The first person's behavior was always high, the third person's behavior was always low, and the second person's behavior was always intermediate; thus, no overlap between the behaviors of the three persons was allowed.

Each actor's three behaviors constitute an intrapersonal scale; the aggregate of nine behaviors constitutes the interpersonal scale. Target behaviors were chosen keeping in mind the fact that each behavior takes on simultaneous values on both scales, and that these values are typically different. For example, a behavior that is high interpersonally may well be low intrapersonally. In all, there are nine possibilities for target behaviors.

The placement of target behaviors in examples was balanced with respect to the two scales, given the constraint that person A's target was always high interpersonal, person B's target was always medium interpersonal, and person C's target was always low interpersonal. To insure that the subjects paid attention to all the data presented to them, a preliminary task was developed for each example. Since one has to look at all three of an actor's behaviors to find her middle score, subjects were asked to pick out the median score for each target actor. This task has the added advantage of having subjects pay special attention to the key reference points for both the interpersonal and intrapersonal distributions.

Design. A manipulation of instructions created three groups. Subjects in the intrapersonal condition were instructed to judge target behaviors by comparing to the actor's previous behavior; subjects in the interpersonal condition were instructed to judge the target behaviors by comparing to the previous behavior of the group; subjects in the unspecified condition were not given instructions as to how to judge the target behaviors. Evaluative judgments were made on a seven point semantic differential scale.

Procedure. The instructions informed the subject that a series of examples would be presented, that each example would contain a summary of an activity such as bowling or competitive sales, that behavior of three individuals would be given for each activity, and that two tasks would need to be performed for each example. The middle-value task was presented first and required the subject to locate the middle score (median) in each actor's distribution of behaviors. The second task was termed the judgment task and required the subject to rate a new behavior from each of the three actors. A new behavior was given for each actor and subjects were to rate it by checking the adjective best completing a stem sentence. It is here that the independent variable was implemented, as the stem sentence was varied by condition. If subjects were placed in the unspecified condition the following stem completion appeared:

Alfred's performance on the fourth afternoon was

[] Very Good
[] Good
[] Fairly Good
[] Nothing Special

[] Rather Bad
[] Bad
[] Very Bad

In the interpersonal and intrapersonal conditions the stem completion task was the same as above except that a relative clause was added to the beginning of the sentence. The interpersonal clause was "compared to the scores of the group." The intrapersonal clause was "compared to his (or her) previous performance."

Results. Table 1.1 lists the interpersonal, intrapersonal, and unspecified means and variances for each target judgment case. Also listed is a p-value for each judgment, which is a measure of the relative weighting of the two standards (e.g., an estimate of the probability of an intrapersonal judgment being made in the unspecified condition), and a model variance estimate based on a combination of the means and variances of the interpersonal and intrapersonal groups (e.g., a prediction of what the variance of the unspecified group should be if the choice hypothesis is true). Finally, an F-ratio is listed for each judgment case. This ratio is composed of the model variance over the variance observed in the unspecified group.

The p-values range from a low of .67 to a high of .97, with the average p-value equal to .81. In all cases, the variance of the unspecified group is considerably greater than the variance in either the interpersonal or intrapersonal groups. In general, these data can be interpreted to suggest that people choose between interpersonal and intrapersonal standards when judging another's behavior. In four of the cases they used the intrapersonal standard outright, rejecting interpersonal comparison completely. In the other twelve, 80% judged intrapersonally and 20% judged interpersonally.

Experiment 2

The purpose of the second study was to determine the influence of the mid-value orienting task utilized in the first study. It is possible that this task may have encouraged the predominant use of the intrapersonal standard in subjects' judgments of behavior. To see if this was the case, a new orienting task was developed. In this task, subjects were asked to order all nine scores in each example form highest to lowest and to write down the second, fifth, and eighth highest ones. Notice that subjects write down the exact same scores in this new "2,5,8 task" as they would in the mid-value task (this is due to the fact that the three distributions in each example do not overlap). By focusing subjects' attention on all nine scores, this new task should have the effect of emphasizing the interpersonal frame of reference more than the intrapersonal frame of reference. Thus, if the orienting task is influencing subsequent judgments of behavior, then judgments following the 2,5,8 task should have lower p-values than judgments following the mid-value task. Conversely, p-values should remain the same if the orienting task has no influence.

Method

Subjects. Sixty-nine University of California undergraduates participated in the experiment in order to fulfill a course requirement. Subjects were run in several sessions.

Materials, Design, and Procedure. Everything was the same as in experiment one except for the new orienting task. At the top of each example subjects were instructed as follows:
2nd, 5th, 8th Task.

Ordering all nine from highest to lowest,please list the 2nd, 5th, and 8th highest scores:

2nd_____ 5th_____ 8th_____

Results. Table 1.2 shows that the p-values have indeed come down. In Experiment 2, p ranges from .25 to .86, with the average p being .48. Thus, subjects clearly judged more interpersonally in the present study than in Experiment 1. However, the effect of the 2,5,8 task seems to be less pronounced than the mid-value task, as p averages about .5. P would have had to average .25 to match the .75 effect of the mid-value task. Table 1.2 also reveals evidence that subjects mixed the two frames of reference. Indeed, in 4 of the 18 cases F reaches significance and allows for a rejection of the choice model.

In sum, Experiment 2 suggests that the mid-value task biases subjects' subsequent judgments toward the intrapersonal frame of reference. Moreover, the alternative 2,5,8 task produces less of a bias, even though subjects search for the same scores as in the mid-value task. In addition, the presence of judgments that combine the two frames of reference suggests the following hypothesis: The orienting task activates, or primes, one of the frames of reference (mid-value primes intrapersonal; 2,5,8 primes interpersonal); however, regardless of task, attributing a score to an individual activates the intrapersonal frame of reference. Thus, when the mid-value task is used, very little consideration of the interpersonal standard will be seen, since it has not become activated. This account does not, of course, explain why 10 - 20% of the subjects in experiment one judged interpersonally.

Experiment 3

The purpose of Experiment 3 was to test the interpretation of the interpersonal instructions. It seems possible that subjects might take interpersonal information into account when making this judgment, even though they have been explicitly instructed to judge intrapersonally. Experiment 3 tests this possibility by introducing a manipulation of the interpersonal scale. If interpersonal information is covertly influencing overt intrapersonal judgments, then it should

make a difference where in the interpersonal distribution the target actor appears. That is, the same target behavior should be rated differently if the actor is at the top of the interpersonal scale than if he is in the middle, since an intrapersonally poor behavior will be interpersonally fair if he is at the top of the distribution, but interpersonally poor if he is in the middle. Two versions of the intrapersonal questionnaire were devised, such that for each example the background and target behaviors for two of the actors were the same between forms, and one actor was different between forms. The different actor was either higher or lower interpersonally than the other two. The point was to see if a target behavior is rated the same when the actor is interpersonally the best of the three (designated actor A), as when he is interpersonally in the middle (designated actor B).

Subjects. 50 University of California undergraduates participated in the study as a part of a course requirement. All subjects were run in individual sessions.

Materials, Design, and Procedure. The materials were as in the previous two studies. In each example, the original background and target behaviors were compressed slightly to make room for a fourth actor's behaviors This was done so as not to extend the range absurdly in several of the examples (for example, a baseball average of .140).

Results. Without question the results do not support the hypothesis of interpersonal pollution. The means of subjects' ratings of target actors common across the two conditions were subjected to t tests. Of the eighteen, only one achieved significance at .05 level (the critical value is $t = 1.69$).

Experiment 4

Experiment 4 tested the idea that reversing the judgment task of the first experiment might lead to more mixing of the frames of reference. Just as interpersonal and intrapersonal norms can be used as judgment standards, they can also be used to generate new behaviors given an evaluative description. So, if I am told that Bill shot a "good" round of golf, I can generate what his score must have been to deserve that description.

Subjects. 56 paid subjects participated as a part of a series of unrelated experiments which were run together.

Materials, Design, Method. Again, the same 9 examples were utilized from experiment 1. The 2,5,8 orienting task was used in place of the mid-value task, because it seems to have a less biasing effect on subsequent judgments. The background behaviors were the same as in Experiments 1 and 2. In place of target behaviors were evaluative

descriptions of behavior on a fourth occasion. These descriptions were chosen to match the target behaviors that were used in Experiments 1 and 2. As in Experiments 1 and 2, three groups of subjects were created -- interpersonal, intrapersonal, and unspecified groups. Subjects in the unspecified condition were given the following judgment task:

Alfred's performance in the fourth game was Nothing Special. He must have shot a score of _____ .

Subjects in the intrapersonal and interpersonal conditions were given the following judgment task with a relative clause added to the beginning of the sentence: "compared to his previous scores," and "compared to the scores of the group," respectively.

Results. The numerical results of Experiment 4 were subjected to the same probability model as the ratings of Experiments 1 and 2. In the High/Low and Low/High cases, p ranges between .59 and .87., with the average p across the six cases being .79. These results look more like experiment 1 than 2. Thus, in the reverse task, people appear to be choosing between frames of reference.

Table 1.1: Results of Experiment 1.

	Intrapersonal		Interpersonal		Unspecified			Model	
Case#	Mean	Var.	Mean	Var.	Mean	Var.	p	Var.	F
1	-1.86	0.48	1.73	0.56	-0.96	3.22	0.75	2.96	0.92
2	-1.05	0.43	1.60	0.61	-0.26	2.93	0.70	1.96	0.67
3	-1.83	0.72	1.45	0.34	-1.43	1.80	0.88	1.81	1.03
4	1.96	0.56	-1.26	0.28	1.30	2.58	0.80	2.20	0.85
5	2.04	0.56	-1.18	0.97	1.78	1.81	0.92	1.40	0.77
6	2.96	0.21	-0.73	0.65	1.87	2.57	0.81	3.28	1.27
7	1.08	0.66	2.76	0.18	1.13	0.85	0.97	0.75	0.88
8	0.78	0.66	2.74	0.19	1.30	1.58	0.73	1.27	0.80
9	0.21	0.58	2.20	0.46	0.87	0.85	0.67	1.45	1.71
10	-2.04	0.32	-0.32	1.04	-1.74	1.38	0.82	0.90	0.65
11	-1.75	0.49	-0.22	0.52	-1.48	0.99	0.82	0.86	0.87
12	0.09	0.25	-2.23	0.54	-0.39	0.88	0.79	1.20	1.36
13	-0.40	0.04	-2.00	0.35	-0.35	0.51	0.84	0.60	1.17

Table 1.2: Results of Experiment 2.

	Intrapersonal		Interpersonal		Unspecified			Model	
Case#	Mean	Var.	Mean	Var.	Mean	Var.	p	Var.	F
1	-2.04	0.41	1.85	0.13	0.09	2.99	0.45	4.02	1.34
2	-1.00	0.36	1.35	0.33	0.61	2.34	0.32	1.55	0.66
3	-2.00	0.70	1.50	0.36	-0.26	1.93	0.50	3.62	1.88*
4	1.52	0.42	-1.39	0.46	-0.09	1.36	0.45	2.56	1.90*
5	2.13	0.29	-1.32	1.06	0.83	1.78	0.62	3.40	1.90*
6	2.43	0.51	-1.53	0.78	0.22	3.36	0.44	4.56	1.36
7	2.04	0.48	1.36	0.23	1.70	0.67	0.50	0.49	0.72
8	1.67	0.51	1.30	0.41	1.39	0.70	0.25	0.48	0.69
9	0.77	0.81	2.44	0.47	1.52	1.62	0.55	1.38	0.85
10	0.65	0.49	2.67	0.22	1.96	1.13	0.35	1.26	1.11
11	0.30	0.30	2.53	0.25	1.61	1.46	0.41	1.49	0.98
12	-1.68	0.94	-0.44	0.36	-0.91	1.36	0.38	0.97	0.71
13	-1.35	0.23	-0.44	0.25	-1.22	0.45	0.86	0.34	0.76
14	-0.14	0.57	-2.00	0.67	-0.96	1.04	0.56	1.50	1.40
15	0.09	0.08	-2.00	0.56	-1.00	1.54	0.48	1.43	0.93
16	0.91	0.45	-1.87	0.48	-0.61	1.25	0.46	2.42	1.94*

-- Model Var. refers to variance predicted by the probability model, e.g., the variance that would be expected if the choice hypothesis is true..
-- F is composed of the model variance (column 9) over the variance observed in the unspecifie condition (column 7).
-- * indicates significance at the .05 level.
-- Five judgment cases were excluded from experiment 1 and three cases were excluded from experiemnt 2 because p could not be estimated.

Section 2: Mental Contamination

Literature Review and Theoretical Analysis (Kahneman and Varey)

Several sources of evidence suggest that intentional control of mental processes is not always as easy as it may appear. In fact, the intention to perform a particular mental operation commonly activates other operations in addition to the specifically intended one. The proliferation of such unintended computations creates a problem of control that is often manifested in slowed responses, in contaminated responses, or in outright errors.

Together with Carol Varey, I am currently engaged in a review of contamination effects in the cognitive and social psychology literatures. We distinguish between two broad categories of effects arising from unintended computations. When responses are made along an ordered scale, the contaminated response reflects a <u>compromise</u> between answers arising from the intended and the unintended processes. In these situations the <u>outcome</u> of the intended process is affected by unintended processing. When the response is a categorical choice, the results of the unintended process provide either <u>conflict</u> with, or <u>support</u> for, the result of the intended process, and crosstalk produces <u>delayed or speeded responses</u>, or errors.

A prototypical example of compromise effects is the phenomenon of anchoring in judgment: the processing of the anchor as a suggested solution to a problem typically leads to a response that is pulled toward the irrelevant and uninformative value. The Stroop effect is a paradigmatic illustration of conflict effects due to an unnecessary mental operation. In the Stroop task, subjects are asked to name the ink-color that a word is written in. Subjects are slower to name the ink-color when the written word is itself a conflicting color word. This effect is not simply a reduced efficiency resulting from performing two processes at once since different words have different effects. The color naming process is slowed down relative to reading a neutral word. And, in fact, a congruent color word results in faster color naming.

Our review explores these and other contamination effects in depth, addressing cognitive variants of Stroop effects, such as the confusions between metaphorical and literal truth, and between truth and validity, as well as manifestations of 'unintended thought' in social perception. In the last year, the grant has supported several experimental research programs in contamination. Karen Jacowitz and I conducted a large study of anchoring effects in judgment; Carol Varey wrote her dissertation on a new source of crosstalk effects; with Anne Treisman and Maria Stone I began a new line of studies on crosstalk between concurrent relational tasks. Further research on crosstalk effects is planned for the extension period.

Crosstalk and Contamination in Cognitive Processes -- Carol Varey

This dissertation investigated the problem of the control of cognitive operations. If a person wishes to perform an operation, A, how effectively can she prevent herself from performing operation B in addition to, or instead, of A? What operations are likely to be performed inadvertently, and why?

The Introduction reviewed several examples in the psychological literature that show that the result of an unintended process can have important consequences on the intended process. The term crosstalk refers to the response timing effects and errors that arise from conflict (or collaboration) between intended and unintended processes. A Theoretical Framework section considered these crosstalk effects in the light of three possible sources for unintended operations: habitual cognitive operations, recently-performed operations, and concurrent operations.

This theoretical framework for conceptualizing crosstalk suggested the possibility of effects not previously investigated in the literature. Two such effects, called computational momentum and stimulus inertia, were investigated in a series of four experiments. The first effect, computational momentum, is the tendency for people to continue to perform a mental operation after it is no longer relevant. Thus, tasks that were intended only to be performed on earlier stimuli are also performed on currently-relevant stimuli, creating crosstalk with the currently relevant task. The second effect, stimulus inertia, reflects the tendency to perform the current operation upon memory traces of stimuli that were processed earlier.

The investigation of computational momentum and stimulus inertia requires an experimental paradigm in which the subjects' task changes frequently. Effects of computational momentum are shown when performance on the intended operation is affected by the answer to the previous operation applied to the current stimulus. Such effects may be evinced by slowed or speeded responses dependent upon the irrelevant answer, or by changes in error rate dependent upon the irrelevant answer. Similarly, effects of stimulus inertia are shown when performance (speed or accuracy) on the intended operation is affected by the answer to the current operation applied to a previous stimulus. Two paradigms allowing frequent changes of task were used: feature verification and "same"-"different" judgments.

Experiments 1 and 2 used a feature-verification paradigm. Subjects were presented with simple visual displays such as three red triangles at the top of the terminal screen, or two blue squares at the left of the screen. In any single display the elements all shared the same color and shape, they were all in the same quadrant on the screen, and there were two, three, four or five elements. Each display was defined by a conjunction of four features (color of elements, shape of elements, number of elements, and screen position of display), with each feature chosen from a set of four possible values. Subjects were presented with a question probing a particular feature value, for example "Blue?" to which they responded by hitting

the key marked "Y" for Yes, or the key marked "N" for No. In Experiment 1, subjects performed the same task for five displays, after which a new question appeared and was in turn applied to five displays, and so on. In Experiment 2, a new question appeared with each display.

An illustration will serve to explain how crosstalk effects can be examined in this paradigm. Suppose that the subject intends to answer the question "Blue?", and that her previous question was "Triangle?" Computational momentum is evinced by differences in the response to "Blue" depending on whether or not the current display shows triangles. Stimulus inertia, in contrast, is shown by differences in the response to "Blue?" according to whether or not the previous display (the target of the "Triangle?" question) was blue or not.

In Experiment 1, there were clear effects of conflict between the computational momentum (CM) answer and the answer to the current (intended) question. These effects were present in both RT and error rates. As predicted, these effects were strongest for the first and second displays following a new question, as shown below:

Table 2.1. Effects of computational momentum on RT for each display in Experiment 1 (n=22).

	Correct answer	CM answer No	CM answer Yes
display 1			
	No	653	667
	Yes	637	594
display 2			
	No	485	493
	Yes	461	451
display 3			
	No	490	496
	Yes	445	450
display 4			
	No	484	490
	Yes	451	446
display 5			
	No	496	496
	Yes	459	446

The answer to the irrelevant stimulus inertia (SI) question also had effects on RT and error rates, although in this case responses to the

current question were faster and more accurate when the SI answer was yes, irrespective of the current answer (see Table 2.2). Although subjects may have computed the irrelevant stimulus inertia answer, an alternative explanation for this result is that when a feature appears in a display it semantically primes the related probe, thus facilitating responses to it.

Table 2.2. Effects of stimulus inertia on RT for display 1, Experiment 1 (n=22).

		SI answer No	SI answer Yes
Current answer	No	659	638
	Yes	628	616

The computational momentum and stimulus inertia effects were markedly larger than the effects of the previous response (see Table 2.3). Also, the faster responses when the previous response was compatible were obtained at the cost of greater errors. In other experiments compatibility with the previous response has been found to influence RT. However, the paradigm of varying questions allows the effects of the previous response response to be unconfounded from the effects of the previous question. It appears that repeating the question may be a more important factor in "response-priming" effects.

Table 2.3. Effects of previous answer on RT for display 1, Experiment 1 (n=22).

		Previous answer No	Previous answer Yes
Current answer	No	658	662
	Yes	620	611

The CM effects in Experiment 1 may have occurred because the questions remained relevant for five trials, or because the question had to be committed to memory. In Experiment 2, these explanations were tested by presenting the question simultaneously with the relevant display, thus eliminating the memory requirement, and changing the question with each display, thus eliminating any benefits to be derived from a processing habit developed over displays. Again, compatibility effects of computational momentum were observed (see Table 2.4).

Table 2.4. Effects of computational momentum on RT, Experiment 2 (n=18).

		CM answer No	CM answer Yes
Current answer	No	878	894
	Yes	852	840

The response to the stimulus inertia question also had an effect on RT, but in this experiment responses were faster and more accurate when the answer to the stimulus inertia question was No (see Table 2.5).

Table 2.5. Effects of stimulus inertia on RT, Experiment 2 (n=18).

		SI answer No	SI answer Yes
Current answer	No	877	896
	Yes	848	852

The remaining experiments used a "Same"-"Different" paradigm to investigate computational momentum. In Experiments 3a and 3b, subjects were first shown one of the questions "Same Color?", "Same Shape?", or "Same Number?". Then they were presented simultaneously with two simple visual displays, one on the left of the screen and one on the right (for example two green crosses on the left, and four white circles on the right). If the displays matched on the probed dimension, subjects responded by pressing a key marked "S" for Same. Otherwise they responded with "D" for Different. As in Experiment 1, subjects responded to five displays for each question.

In this paradigm, evidence for computational momentum is shown by an effect of the CM answer (say, shape same or different) on the current answer (say, color same or different). Table 2.6 shows that CM effects are large and appear to be maintained across all five displays.

Table 2.6. Effects of computational momentum on RT for each display, Experiment 3a (n=20).

		CM answer Diff	CM answer Same	
relevant similarity	Diff	783	836	stim 1
	Same	709	686	
relevant similarity	Diff	609	605	stim 2
	Same	588	552	
relevant similarity	Diff	602	620	stim 3
	Same	567	550	
relevant similarity	Diff	608	622	stim 4
	Same	567	539	
relevant similarity	Diff	629	646	stim 5
	Same	590	566	

It was necessary to test whether these results were due to computational momentum, or were an artifact arising from a tendency for subjects to process <u>all</u> similarity dimensions, regardless of whether the dimension was recently probed. This was investigated in Experiment 3a by comparing the effects of irrelevant shape similarity for cases in which shape was the previously-probed dimension, with cases in which it was not. In Experiment 3b only the color and number probes were used. This allows us to see whether there is any effect of crosstalk from a dimension that is never probed. As table I-7 shows, the compatibility effects of irrelevant shape similarity are much larger when shape was the previous question (i.e. shape is the CM dimension).

Table 2.7. Effects of irrelevant shape answer on RTs in Experiments 3a and 3b.
Columns (1) and (2) are from Experiment 3a (n = 20); Column (3) is from Experiment 3b (n = 20).

		(1) irrelevant shape is CM dimension		(2) irrelevant shape is not CM dimension		(3) irrelevant shape is never probed	
		Shape Diff	Shape Same	Shape Diff	Shape Same	Shape Diff	Shape Same
Color relevant:							
Color	Diff	566	591	569	620	612	628
Color	Same	513	502	515	516	548	559
Number relevant:							
Number	Diff	702	749	696	696	738	722
Number	Same	696	608	669	611	723	666
means:							
	Diff	634	670	633	658	675	675
	Same	604	555	592	563	636	613

Experiment 4 extended the feature version of the "Same"-"Different" paradigm to investigate cross-modal crosstalk. Subjects were given "Same Tone?" or "Same Color" as a probe, then the first color was presented accompanied by a tone, followed by the second color-tone pair. As in Experiment 3a, computational momentum was examined as a possible modifier of concurrent crosstalk effects. Results showed that the effects of irrelevant similarity were much larger when the irrelevant dimension was probed in the previous question (see Table 2.8). Again, conflict with the computational momentum answer led to slower responses than responses supported by the computational momentum answer.

In summary, all the experiments showed that the result of the computational momentum process affected the speed and accuracy of responses to the relevant question. The effect was observed in both feature-verification and "same"-"different" paradigms. Crosstalk occurred when the CM question probed a different modality from the currently-relevant question, as well as when both questions referred to a visual dimension. Experiment 2 showed that computational

momentum effects do not appear solely as a result of a set of repeated applications of a particular operation, since a single trial will suffice. Nor is committing the task to memory prior to the relevant trials a necessary condition for computational momentum, since the effect is still evident when the task and the stimulus are displayed together. Thus it appears that even after a single execution of a task people have a tendency to repeat the same operation, and the results of the unnecessary operation contaminate the intended process. Future research is planned to investigate these effects further.

Table 2.8. Effects of irrelevant-modality answer on RT across all displays, Experiment 4 (n=19).

	(1) Other dimension probed in previous trial		(2) Same dimension probed in previous trial	
	irrelevant answer		irrelevant answer	
relevant dimension	Diff	Same	Diff	Same
Tone:				
Tone Diff	382	425	400	413
Tone Same	400	369	377	361
Color:				
Color Diff	376	358	356	344
Color Same	325	318	338	302

Contamination effects in comparison. (Kahneman, Treisman and Stone)

Carol Varey discussed crosstalk effects arising from the performance of unintended operations on the stimuli on which the intended operations are performed, or else of performing the current operation on a memory trace of the stimuli that were processed earlier. It is also possible to perform the intended operations on concurrent stimuli which should be ignored, because they are never relevant to the task. Two pilot experiments investigated the conditions under which this becomes a problem in the context of comparisons.

In the first experiment, subjects were presented with four objects on the screen. The objects were two vertical lines or two digits in the

middle, flanked by two tilted lines. The subjects' task was to disregard the middle objects, and to press the right key if the right tilted line was shorter, and the left key if the left tilted line was shorter. There were six conditions in this experiment, three with the vertical lines in the middle, and three with digits in the middle. When the lines were in the middle, the shorter of the two vertical lines could appear on the same side as the shorter tilted line (consistent condition), or on the opposite side (inconsistent condition); in the control condition, both middle lines were the same length. When the digits were in the middle, the smaller (numerically) of the two digits could appear on the same side as the shorter tilted line (consistent condition), or on the opposite side (inconsistent condition); in the control condition, both digits were the same. In the second experiment, the two flanking objects were digits, and the subjects' task was to press the key corresponding to the digit that was numerically smaller. This experiment had the same six conditions determined by the nature of the middle two objects. If subjects unintentionally compared the two middle objects, they should be faster in the consistent condition and slower in the inconsistent condition. The results of the pilot experiments are presented in the following two tables

Table 2.9. Contamination effects on line comparison task (mean reaction times, n=14)

distractor stimuli	type of condition		
	control	consis	inconsis
lines	708	733	753
digits	702	706	746

Table 2.10. Contamination effects on digit comparison task (mean reaction times, n=15)

distractor	type of condition		
	control	consis	inconsis
lines	522	523	525
digits	521	529	530

In the first experiment, there is significant effect of the compatibility of the digits (40 msec., $t(13)=2.92, p<0.1$), but not of the lines (20 msec., $t(13)=0.87$). The effect obtained with the digits is remarkable because of the strong subjective impression that the digits are not processed at all, and are indeed virtually invisible. In the second experiment, there is no interference from the lines, or from the digits. The reaction times were also very much faster in the second experiment. The absence of a line effect in Experiment 2

is easily explained: as in the Stroop case, a slow task has little effect on a fast one. However, the total lack of effect of the focally presented irrelevant lines in Experiment 1 and of the digits in Experiment 2 cannot be explained in the same fashion. One possibility, which is compatible with early research by Treisman and Fearnley (1969) is that the intentional processing of the peripheral items prevents the same kind of processing from being applied to other stimuli.

The interesting result of these experiments, of course, is the positive effect that was observed from digits on the line length task. Here an intention to respond to the shorter of two digits spilled over into a tendency to respond to the smaller of two digits. We hope to follow up this result and to use the technique in an effort to map the representation of various tasks that involve the detection of relations between stimuli.

Section 3: Anchoring Effects
Kahneman and Jacowitz

The phenomenon of anchoring occurs when some initial value exists that a subject uses as a starting point for determining a response to a stimulus. Most often in the research to date, the anchor value has been a number that appears somewhere in the question or in the introduction or instructions. Then, subjects can adjust this value in the direction that they feel is appropriate in order to generate their actual response. In general, researchers have found that subjects do not make sufficient adjustments, so their final judgment is "anchored" to the initial value.

Many researchers have studied anchoring effects on judgment tasks and those factors that make them more or less likely to occur. Markovsky (1988) proposes three conditions for anchoring to occur: 1) the judgment is indeterminate, 2) an anchor exists, and 3) the anchor is salient. In addition, a potential anchor is more likely to be used as such if it is in a format that is compatible with the response scale (Schkade and Johnson, 1989).

In some cases, factors that were predicted to reduce anchoring effects, such as uncertainty (Cervone and Peake, 1986), high time pressure and low evaluation apprehension (Kruglanski and Freund, 1983), did so. In other cases, factors predicted to reduce anchoring, such as expertise (Northcroft and Neale, 1987), increased familiarity with the situation (Wright and Anderson, 1989), high levels of concern and vivid imagery (Plous, 1989), were not found to do so.

Another factor that might reduce anchoring effects is the degree of knowledge that subjects have about a topic and their confidence in their judgments. Although this has been suggested (e.g. Plous, 1989), no empirical support has demonstrated that susceptibility to anchoring is inversely related to confidence. In this study, we tried to provide direct empirical support for this relationship.

In order to test whether high confidence reduces anchoring effects, we needed to have a method for measuring anchoring. There

are certain logical constraints on how to measure anchoring. For instance, at least two different anchors are needed for each question, as well as an unanchored group in order to compare the distributions of responses with and without anchors. The second purpose of this research is to provide an index that represents a measurement of the amount of anchoring in the responses to numerical judgments. The index value is determined by finding the difference between the means of groups exposed to high and low anchors. This difference is then divided by the difference between the anchor values. The index represents a measurement of the amount of motion toward the anchor values. For example, if the difference between the means is the same as the difference between the anchor values, that would indicate perfect anchoring and the index value would be one. If there is no difference between the means of the high and low anchor groups, then apparently the different anchors had no effect. In such a case, the index will equal zero which means that no anchoring has occurred. As the difference between the means increases, the high and low anchors are having more of an effect on the distributions. As a result, the index value will increase.

In order to be able to determine what would be appropriate high and low anchor values, we first obtained a distribution of unanchored responses to each of our 15 questions. The anchors that we used for the experimental groups were the 15th and 85th percentile responses from the unanchored distribution. Because the subjects in the pretest and experimental groups were taken from the same population, we would expect the distributions to be similar if the anchors had no effect. However, if the anchors did have an effect, we would expect the distributions to shift so that the distribution of responses in the high (low) anchor condition would in general be higher (lower) than in the unanchored condition. We would also predict that highly confident subjects would be less affected by the anchors than less confident subjects.

Method

Subjects were 156 students at the University of California, Berkeley. They completed the questionnaire as partial fulfillment of a course requirement in an introductory psychology class.

Subjects were asked to give their best estimates in response to 15 questions. Then, they were asked to rate their confidence in their answer on a ten point scale on which 0 was labeled "not at all confident," 5 was labeled "moderately confident," and 10 was labeled "extremely confident." Questions included some measurements such as the height of Mount Everest and some quantities such as the number of nations that are members of the United Nations.

Pretest subjects (N=53) were asked the questions directly. Anchor values for each question were chosen as the 15th and 85th percentile responses from the distribution of the pretest subjects' responses.

Experimental subjects (N=103) answered pairs of questions. The first question asked whether the quantity in question was greater or less than an anchor value. The second question was identical to the

pretest questions which asked for a specific answer. There were two versions of the questionnaire, each with half high anchors and half low anchors.

Results

In order to provide a measurement of anchoring, an index of motion toward the anchor was developed. The index for each question was defined to be the distance between the medians obtained with the high and low anchors divided by the distance between the high and low anchor values. An index value of 0 would indicate that no motion toward the anchor occurred because the two medians are identical. Greater values of the index indicate a higher degree of anchoring effects because the medians are farther apart (see Table 3.1).

To test the hypothesis that the degree of anchoring is inversely proportional to the level of confidence, the correlations between the index values and the mean and median confidences were calculated separately for the unanchored and anchored groups. For the unanchored groups, the correlation with the mean confidence was $r=-.675$ ($r^2=.455$) and the correlation with the median confidence was $r=-.741$ ($r^2=.549$). For the anchored groups the relationship was even stronger. The correlation with the mean confidence was $r=-.818$ ($r^2=.669$) and the correlation with the median confidence was $r=-.840$ ($r^2=.705$).

To further examine this relationship, low confidence subjects were separated from high confidence subjects for each question using a median split and separate index values were calculated. For all but one question, the index value is lower for the high confidence than low confidence subjects (see Table 3.1). Thus, highly confident subjects were less affected by the anchors than were less confident subjects.

To test whether the distributions of responses were significantly affected by the high and low anchor values, Mann-Whitney tests were performed for each question. All of the differences were highly significant (see Table 3.2).

ıoring Results:

le ·3.1

hors:		Medians:		Index:		
Anch 1	Anch 2	Anch 1	Anch 2	Overall	Lo Conf	Hi Conf
70	2000	300	1500	.62176	.62176	.72539
45500	2000	42550	8000	.79425	1.03448	.50575
1000	50	500	100	.42105	.44895	.24211
1500	6000	2600	4000	.31111	.66667	.11111
65	550	100	400	.61856	.82474	.30928
14	127	26	100	.65487	.66372	.61947
130	25	95	50	.42857	.48571	.38095
5000000	200000	5050000.5	600000	.92708	.98958	.39583
1850	1920	1870	1900	.42857	.42857	.40000
50000	100	40000	1000	.78156	.78657	.68136
30	7	20	10	.43478	.43478	.17391
80	20	60	40	.33333	.41667	.21667
10	85	20	40	.26667	.26667	.13333
20	100	30	50	.25000	.56250	.06250
17	7	16	16	0	.05000	0

le 3.2

ın-Whitney Tests

N(A2)	N(A1)	Z	p
51	50	7.315	0
50	51	-7.512	0
51	51	-7.037	0
50	51	5.142	0
51	51	6.274	0
52	50	6.366	0
51	50	-5.284	0
51	50	-6.942	0
51	47	5.650	0
51	50	-8.124	0
51	51	-6.027	0
51	50	-5.420	0
49	51	3.561	.0004
50	51	3.804	.0002
49	50	-2.806	.0052

Section 4: Topic and Referent in Perceptual Comparisons
Research conducted by Maria Stone

Human thought is selective. This claim is not controversial as long as the thought involves only one object to the exclusion of others. Picking out a single figure from a background or concentrating on a specific object or person in order to retrieve their characteristics from memory are such uncontroversial cases. If linguistic description is warranted, the subject of the sentence will frequently correspond to this selected "topic" of thought.

However, there are many situations when human thought appears to be about not just one, but exactly two objects and a relationship between them. One example is comparisons. In language, different roles are assigned to the two objects involved. One of them becomes the subject (topic) of a sentence, and the other becomes the object, or referent. What is the cognitive significance of this assignment of roles? One possibility is that the thought is about the relationship and/or difference between the objects, and that the assignment of roles arises only when the thought is processed for communication. The other is that the thought is not about the difference, but about one of the objects and its relationship to the other object. In this case, the distinction between the topic and the referent is cognitive as well as linguistic. This research explores the cognitive consequences of directional comparisons.

Maria Stone's previous research examined how the topic can be designated in linguistically neutral comparisons. The experiments described in an earlier report explored the link between attention and the selection of the topic of comparison. This year, the focus of research was on distinguishing the kind of processing the topic and the referent receive in perceptual comparisons. Two aspects of this distinction have been proposed.

1. The topic is said to "control the agenda" for comparison; e.g., the features of the topic get mapped onto the features of the referent, but not vice versa. This should have several empirical consequences.

(a). When the topic has more unique features than the referent, it appears more different from the referent than when the referent has more unique features than the topic. This asymmetry was studied by Tversky (1977) and Agostinelli et al. (1986). It was also utilized in the six experiments described in a previous report, which studied the factors that determine the topic of comparison.

(b). For some stimuli, there is a specific natural order in which the features of an item are encoded (eg., letters in words). When two such items are compared directionally, the order in which the features will be checked off should correspond to the order of the features in the topic item.

(c) If the common features group together (due to proximity or similarity) in the topic, but not in the referent, finding them should be easier than when they group together in the referent, but not in the topic.

2. In the process of comparison, the topic is encoded relatively, whereas the referent is encoded absolutely. The results of this encoding should be noticeable when:

(a). The topic or the referent are repeated in a new comparison.

(b). In the memory for the topic and for the referent.

Overview of the new experiments:

A). Demonstrating that the topic "controls the agenda" of comparison:

Several experiments were conducted to demonstrate that the order in which the features of the two objects are compared is determined by the order of features in the topic object. Five-letter nonsense strings of consonants were used. One of the strings was designated as the topic of comparison using some of the manipulations that were effective in the previously reported experiments. The subjects' task was to write down the letters that the strings had in common. The strings were randomly generated, and always had three letters in common and two unique letters each. The order in which the common letters appeared in the two strings was randomly determined, and was often (but not always) different. Subjects were expected to report the common letters in *the order in which* they appear in the topic string.

In the first experiment, the first string was presented for 2000 msec., then a mask of "XXXX" was presented for 170 msec, then a long interval (1000 msec), and, finally, the second string was presented for 2000 msec. The results of previous experiments suggest that the first string should become the topic of comparison in this situation, i.e., the subjects will report the common letters in the order in which they appear in the first string. The results confirm this prediction--subjects were more likely to report the common letters in the order in which they appear in the first string than in the order in which they appear in the second string. The entire experiment consisted of 20 trials, and on average, on 8.2 trials the order of the reported letters was consistent with the order of common letters in the first string, compared with only 4.3 trials for the order consistent with the second string.

A second manipulation was designed to assign the role of topic to the item shown last on a trial. Two strings were shown on each trial, one in capitals and one in lower case. The strings remained on the screen for the duration of the trial. A third string, added 2000 msec later, could be either in capital or in small letters. The subjects' task was to compare the two strings in the same case. Previous results suggested that in this situation the third string would be the topic of comparison. As before, the hypothesis is that the order in which the common letters appear in the report should

correspond to their positions in the topic string. This prediction was confirmed. This experiment also consisted of 20 trials, and the order of reported letters was consistent with the order of the common letters in the last string on 7.3 trials, compared with 3.4 trials for the order consistent with the string presented earlier, (n=12).

In a third experiment, only one string appeared initially on the screen, followed 2000 msec later by another string. The two strings remained on the screen together for another 1000 msec. The order of the reported letters was consistent with the order in the first letters on 4.9 trials, and with the order of letters in the second string on 4.8 trials (n=36). It appears that in this experiment, subjects were not consistently selecting the same string as the topic.

One problem with this paradigm is that the task is very difficult, and performance therefore strategic, rather than spontaneous and automatic. Exposure parameters had to be adjusted to allow adequate performance, which also meant that the strings stayed on the screen long enough to allow multiple eye movements, and possibly several checks and rechecks of each string. The obtained results may be due to subjects' strategies, rather than to the spontaneous allocation of the role of a topic to one of the objects. New experiments are planned that will use three-letter nonsense strings with only two letters in common, thus making the task easier. The timing parameters will be changed to speed up the presentation. Both the hypothesis about the order in which the features are compared (b) and the hypothesis about the role of grouping (c) will be tested, using the new stimuli.

B). Demonstrating that the topic is encoded relative to the referent, and that the referent is not encoded on the same way.

The present analysis implies a difference between the coding that the topic and the referent are assigned as the result of their comparison. The topic is assumed to be encoded relative to the referent, whereas the referent is encoded absolutely. A new paradigm was designed to demonstrate this. On each trial, subjects were presented with two letters or two digits. One of the items was flashing, and thereby designated as topic. Subjects had to decide whether the flashing item was smaller (for digits) or earlier in the alphabet (for letters). On some trials, either the flashing or the stationary item was repeated from the previous trial. The item could be associated with the same response as on the previous trial, or with the opposite response. Since the topic (flashing item) is encoded relatively, its repetition with the repeated response should be significantly faster than its repetition with the opposite response. Since the referent (stationary item) is encoded absolutely, there shoul be no difference between repeating the referent with the same or with a different response. Results are presented in the following two tables.

Table 4.1: Effects of stimulus and response repetition in the letter comparison experiment.
mean response times for each condition (n=15)

type of perceptual repetition

response	none	top-top	ref-ref	ref-top	top-ref
same	1066	1024	1101	1186	1002
diff	1075	1183	1043	1133	1048

Table 4.2: Effects of stimulus and response repetition in the digit comparison experiment.
mean response times for each condition (n=17)

type of perceptual repetition

response	none	top-top	ref-ref	ref-top	top-ref
same	763	751	762	802	774
diff	794	841	802	774	793

No general benefit of perceptual repetition was observed for either letters or digits. In fact, conditions with no perceptual repetition were faster both for digits ($t(16)=2.93$, $p < 0.01$) and for letters($t(14)=1.99$, $p < 0.10$). For digits, but not for letters, a small benefit of response repetition was present ($t(16)=2.95$, $p < 0.01$). In both experiments, subjects are slower when the topic (flashing) item is repeated with a new response than when the topic (flashing) item is repeated with the old (repeated) response ($t(16)=4.5$, $p < 0.005$ for digits; $t(14)=2.83$, $p < 0.01$ for letters). The effect of repeating the topic is smaller (for digits) or apparently absent (for letters). The difference between the effects of repeating topic or referent is significant both for digits ($t(16)=2.44$, $p < 0.025$) and for letters ($t(14)=3.74$, $p < 0.005$)

The results so far support the hypothesis that the topic is encoded relatively (as being smaller or larger, earlier or later in the alphabet), whereas the stationary (referent) item is not encoded in this fashion. When the relative codes assigned to a topic on two successive trials are in conflict, interference occurs. Since the referent is not encoded relatively, no interference is observed when a new response is paired with a repeated referent.

Another paradigm to be tried out soon will explore the effects of directional comparison on memory for the topic and for the referent.

The following hypothesis will be tested. Since the topic is encoded relatively, the memory for it should be substantially better in the context in which it originally appeared during the comparison (that is, the referent should be a good cue for the topic). If the referent is encoded absolutely, however, memory for it should be context-independent (the topic should not provide a good cue for the referent). The following experiment will test this prediction. On each trial, subjects will be asked first to make a comparison of two lines, and press a key for the one that is longer (shorter). After a brief interval, two new lines will be presented, one of which will be marked by an arrow. The subjects' task will be to recall if the marked line is exactly the same length as one of the lines on the previous trial. The target line could be either the topic or the referent of the preceding comparison (determined by the line associated with the correct response). The prediction is that the effect of repeating the context will be greater for the topic than for the referent.

Section 5: Reference Effects in Choice

Reference Effects in Consumer Choice - O'Curry, Lovallo, & Kahneman

Two experiments were carried out to test the idea that consumers may use the good they usually buy as a referent to evaluate alternatives. While some research on pricing has made use of the notion of loss aversion to explain brand choice, the idea of a reference point in consumer choice has generally been limited to the domain of money. In this research, loss aversion was extended to the domain of product quality to explain asymmetric price competition between national and private label brands.

The key assumption behind this research is that consumers develop a reference point for both price and quality, and that alternatives to the referent are compared on both dimensions. Thus, a consumer who normally buys a high quality item may not respond to price decreases of low quality items because switching involves a gain of money at the cost of a loss in quality (assuming that price and perceived quality are positively correlated). However, consumers who normally buy lower quality goods may well switch up when higher quality goods price deal, because they face a loss of money in exchange for a gain in quality - a standard buying transaction.

Study 1. Loss Aversion for Quality

While loss aversion has been experimentally demonstrated for various goods (Kahneman, Knetsch, & Thaler, 1990) and disadvantages on non-monetary attributes have been shown to loom larger than advantages in a variety of hypothetical choice situations (Tversky & Kahneman, 1990), loss aversion for a single dimension of a consumer good has not previously been demonstrated in a real choice situation.

The buying-selling discrepancy which characterizes the "endowment effect" lends itself to demonstrating loss aversion for quality.

If subjects are loss averse for quality, they should demand more compensation to switch to a lower quality good from a high quality good than they would be willing to pay to acquire the same good when no loss of quality is involved.

Method

Seventy-six psychology undergraduates at U.C. Berkeley participated in this study. Subjects were run in small groups and received course credit for their participation.

Chocolate was chosen as the good for this experiment because distinct quality levels exist within the product category and undergraduates have experience buying it. Toblerone Chocolate was used as the high quality good. Toblerone is well known, has a high quality reputation, and commands a premium price of $1.79 for a 100 gram bar. The low quality chocolate was "Chocolaty" chocolate, a chocolate flavored bar obtained at Newberry's. While this brand is less familiar, the packaging looks cheap and the label clearly states that it is chocolate flavored, rather than real chocolate. Chocolaty bars were priced at three bars for $1.00, for the 3 ounce size. Price information was masked and subjects who asked about price were told that this question would be answered at the end of the experiment.

As in studies of the endowment effect, subjects were assigned to the roles of buyers, choosers or sellers. To avoid social comparison, only one condition was run within a group. Sellers received a Toblerone bar and a form asking them to indicate at what price they would be willing to return their Toblerone in exchange for a Chocolaty bar plus the amount of money. Choosers were given a form that informed them they had a choice between a Toblerone bar or a Chocolaty bar plus an amount of money. Buyers were given a Chocolaty bar and asked how much they would be willing to pay to exchange their Chocolaty bar for a Toblerone bar. In all cases, the amounts of money were listed in 10¢ intervals from $2.50 to 0. Subjects were told to treat each row as a separate decision. To emphasize the importance of indicating their true values, they were told that an amount of money would be announced later and that whatever they had decided for that amount would be executed. Instructions were both written and oral, with care taken to be sure that subjects properly understood the task.

Results and Discussion

Medians were computed for each condition and were as follows: sellers, $1.00, choosers, $.70, and buyers, $.50. The pattern of results indicates a 2:1 buying-selling discrepancy, with choosers closer to the buyers than to the sellers, the same basic pattern found in standard endowment effect experiments. Mann-Whitney analysis shows that medians for the buyers and choosers are not significantly different, $z = .92$, while sellers are significantly different from both buyers, $z = 3.82$, and choosers, $z = 2.85$. These results support the idea that loss aversion for product quality does exist and that loss aversion for money is not a major factor in the discrepancy between buying and selling prices.

Study 2. Differential Response to Price Changes

The second study looks at the case in which prices rise or fall together. When prices fall, the consumer can maintain the current level of quality and pocket the difference between the regular and sale price as a subsidy. Alternatively, a higher level of quality can be obtained with the regular expenditure. The price decrease may act as a windfall (Arkes, Joyner, Nash, Pezzo, Christensen, Schweigert, Boehm, Siegal-Jacobs, & Stone, 1990) and lead the consumer to improve quality. When prices increase, the consumer must increase expenditure to maintain quality or accept a loss in quality to maintain spending at the current level. Without the windfall gain provided by a price decrease, consumers will choose the option that hurts the least.

Method

The experiment was conducted in a classroom setting, with MBA students as subjects. Only those subjects who identified themselves as regular purchasers of beer were used. Beer was chosen because the category includes a wide variety of brands with a positive price-quality relationship. Eighteen subjects met the criterion of being regular purchasers of beer.

Subjects first saw a list of 25 different beers varying in quality, listed with the current retail price from a large grocery chain. They were asked to indicate which beer they would be most likely to buy at the prices listed, in order to establish a reference point for each brand.

Subjects were then asked to indicate their choices in two different scenarios. In one, they were to imagine that they were in a specialty store where prices were 30% higher than regular grocery store prices, and that it would be terribly inconvenient for them to go to their regular store. In the other, they were asked to imagine that their regular store was having a one time promotion in which the prices of all beers were lowered by 30%. For both scenarios, subjects were instructed to pick the one beer they would be most likely to purchase. Finally, subjects were asked to rate the quality of each beer on a 1 - 9 scale, with a "don't know" option for unfamiliar beers.

Results and Discussion

In the lowered price condition, 16 of the 18 subjects switched to a higher quality beer. Of these, 13 switched to a beer which they considered to be in the highest quality category, and 2 switched to the second highest. Both of these subjects rated only a single beer higher than the one they switched to. The subjects who did not switch were already regular purchasers of beer that they considered to be at the highest quality level of those listed.

In the higher price condition, 8 of 18 subjects refused to switch at all. Two subjects switched to a cheaper beer that they rated as equal in quality to their regular beer. The remaining 8 subjects did switch down in quality.

The mean price of the beer chosen in each condition was another measure of interest, because it demonstrates the importance of reference level of expenditure. In the regular price condition, the mean was 4.16, in the lowered price condition, 4.20, and in the raised price condition, 4.64. The difference between the regular and lowered price condition is not significant, although subjects could have stayed with their regular beers and saved money. However, the difference between regular and raised price conditions is highly significant, $t(17) = 3.00$, $p = .008$. Apparently subjects who felt quality to be more important than money were willing to spend significantly more than the reference expenditure to maintain it. Table 5.1 summarizes the results.

Table 5.1 Results, Experiment 2

REGULAR PRICE CONDITION	
Mean quality rating (on a 9 point scale)	6.05
Mean price paid	4.16
LOWERED PRICE CONDITION	
# of subjects who switch to higher quality	16
(4 were already at ceiling, but switched beers)	
# of subjects who refuse to switch	2
(Both considered their regular beer to be the highest quality)	
Mean increase in quality from regular beer	1.89
Mean price paid	4.20
RAISED PRICE CONDITION	
# of subjects who switch to lower quality	8
# of subjects who switched to cheaper beer, same quality	2
# of subjects who refused to switch	8
Mean decrease in quality from regular beer	-.83
Mean price paid	4.64

Conclusions

The experiments provide support for the idea that consumers use regularly purchased goods as reference points for the evaluation of alternatives. Loss aversion for quality was demonstrated, as well asymmetric response to price increases and decreases. Further work in this area will attempt to illustrate the importance of mental accounting (Thaler, 1985) in response to price changes.

Order Effects and Comparison in Choice -- Kahneman and O'Curry in collaboration with Sherman and Bell

Several studies were carried out which extend the Sherman and Kahneman study reported last year. The unifying theme was an attempt to understand the different roles played by order of alternatives and reference points in choice.

The first study tested the hypothesis that having a personal referent for a comparison would override verbal manipulations such as endowing subjecs with an alternative. Results from a preliminary study reported last year showed effects of endowment and primacy for vacations and courses, but not apartments. Because all subjects have some sort of personal referent for the place they live, it seemed quite plausible that personal reference points might have interfered with the experimental manipulation of reference state. The original experiment was extended to a range of stimuli, some of which most subjects have extensive experience with (TV sets, restaurant meals, laundromats) and some of which very few subjects have experience with (condo rentals, laptop computers). In addition to rating their preference on a 13 point scale, subjects indicated which of the items they either owned or were familiar with. Although the results showed a general effect of primacy and endowment, ownership or familiarity seemed to play no role in evaluating alternatives.

Following this failure, we concentrated on understanding the conditions under which primacy effects were likely to occur, and to disentangle effects of endowment and primacy. Studies were run both in Berkeley and at Indiana. Results from Sherman and his graduate students suggested the presence of both endowment and primacy effects, which were presumed to combine additively.

The minimum conditions for a primacy effect seemed to be that the first option must be acceptable and possess unique features that will be noticed as missing in the second option. We speculated that perhaps a "change-of-standard" was responsible - perhaps the first option served as the standard of comparison for the second option, which would in turn serve as a standard of comparison for a third option. This idea yielded the following prediction: if a pair of equally attractive options is preceded by an inferior option, the first option of the pair should be judged more attractive than the second - a primacy effect. If the first option is decidedly superior to the pair of options, the first option of the pair should seem very unattractive, but the last option should be judged more attractive than the second option - a recency effect. This was tested with a simple design. Subjects were given descriptions of three items in several categories - apartments, cars, partners for a class project, blind dates, restaurants, and vacation trips. The first item was either clearly superior or inferior to the other two, which were matched in attractiveness. The task was to indicate which items were the best and worst in each category. In the case where the first item was inferior, the second item should have been judged best, while in

the case where the first item was superior, the second item should have been judged worst. Results were inconclusive - recency effects generally seemed to be more prevalent than primacy, but there was a great deal of variability between items. We plan to rerun this experiment with richer descriptions of stimuli this fall, using a measure of rated attractiveness rather than the "best"/"worst" measure originally used.

Two new directions of research have their roots in the collaboration with Sherman. First is an investigation into the limits of simulation. It occurred to us that perhaps the reason that primacy seemed to have the same effect as endowment was that subjects were not simulating the pain of loss present in standard endowment effect experiments, which involve real goods rather than hypothetical situations. We have run a study which is identical to standard endowment experiments, except that instead of receiving a good, subjects are asked to imagine that they have been given a good. Two conditions were run, using pens and restaurant meals as stimuli. Results indicate no endowment effect for the pen and a reduced endowment effect for the restaurant meal. The presence of the effect for the restaurant meal suggests two alternatives - either the magnitude or the type of good could make a difference to subjects' ability to simulate possession and loss of a good. We will run two more conditions of this experiment, with a low value "frivolous" good and a higher value practical good.

The second direction is an attempt to disentangle the constructs of loss aversion and status quo bias (Samuelson & Zeckhauser, 1988). While the constructs have been used almost interchangeably in the literature, we believe that loss aversion applies only to situations in which the pain of loss is felt. In contrast, status quo bias may apply more widely, taking the form of a rule-like approach to choice where the rule is, "Stay with what you already have, unless something clearly better comes along." Status quo bias is also evident in the asymmetric regret associated with acts of omission and commision, where loss aversion almost certainly plays no role. While several ideas have been discussed, the one that we will investigate next is the effect of similarity of alternatives on loss aversion and status quo bias. We expect to see little loss aversion for very similar items, as measured by willingness-to-accept measures of value, while status quo bias may manifest itself as a general reluctance to trade.

References

Arkes, H. L., Joyner, C. A., Nash, J. G., Pezzo, M., Christensen, C., Schweigert, W., Boehm, L. E., Siegal-Jacobs, K., & Stone, E. (1990). The psychology of windfall profits. Paper presented at the annual meeting of the Psychonomic Society, New Orleans, LA.

Campbell, J.D., Fairey, P.J., & Fehr, B. (1986). Better than me or better than thee? Reactions to intrapersonal and interpersonal feedback. Journal of Personality, 54, 479-493.

Cervone, D. and Peake, P. (1986). Anchoring, efficacy and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. Journal of Personality and Social Psychology, 50, 492-501.

Farkas, A.J. (1991). Cognitive algebra of interpersonal unfairness. In N.H. Anderson (Ed), Contributions to information integration theory. (Vol II: pp. 43-97). Hillsdale NJ: Lawrence Erlbaum Associates.

Hertzman, M., & Festinger, L. (1940). Shifts in explicit goals in a level of aspiration experiment. Journal of Experimental Psychology, 27, 439-452.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1990). Experimental testsof the endowment effect and the Coase theorem. Journal of Political Economy, 98, 1325-1348.

Kahneman D., & Miller, D.T. (1986). Norm Theory: Comparing reality to its alternatives. Psychological Review, 93, 136-153.

Kruglanski, A. and Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. Journal of Experimental Social Psychology, 19, 448-468.

Levine, J.M., & Green, S.M. (1984). Aquisition of relative performance information: the roles of intrapersonal and interpersonal comparison. Personality and Social Psychology Bulletin, 10, 385-393.

Markovsky, B. (1988). Anchoring justice. Social Psychology Quarterly, 51, 213-224.

Northcroft, G. and Neale, M. (1987). Experts, amateurs and real estate: An anchoring-and-adjustment perspective on property pricing decisions. Organizational Behavior and Human Decision Process, 39, 84-97.

Plous, S. (1989). Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war. Journal of Applied Social Psychology, 19, 67-91.

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. Journal of Risk and Uncertainty, 1, 7-59.

Schkade, D. and Johnson, E. (1989). Cognitive processes in preference reversals. Organizational Behavior and Human Decision Process, 44, 203-231.

Scul, Y. & Szyf, M. (1991). Implicit theories of satisfaction: concurrentinfluences of temporal and interpersonal standards. Unpublished manuscript: Hebrew University, Jerusalem.

Thaler, R. (1985). Mental accounting and consumer choice. Marketing Science, 4, 199-214.

Tversky, A. & Kahneman, D. (1991, in press). Loss aversion in riskless choice: A reference dependent model. Quarterly Journal of Economics.

Wright, W. and Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. Organizational Behavior and Human Decision Process, 44, 68-82.